

AD-A119 370

STANFORD UNIV CA DEPT OF STATISTICS F/8 12/1
ON THE ASYMPTOTIC DISTRIBUTION OF THE SIZE OF A STOCHASTIC EPID--ETC(U)
MAY 82 T SELLKE
UNCLASSIFIED TR-17 N00014-77-C-0306
NI

1-1
1-1
1-1

1-1

END
DATE
FILMED
10 82
DTIC

(12)

ON THE ASYMPTOTIC DISTRIBUTION OF THE
SIZE OF A STOCHASTIC EPIDEMIC

BY

THOMAS SELLKE

TECHNICAL REPORT NO. 17
MAY 1982

AD A119370

PREPARED UNDER CONTRACT
N00014-77-C-0306 (NR-042-373)
FOR THE OFFICE OF NAVAL RESEARCH

This document has been approved
for public release and sale; its
distribution is unlimited.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA



DTIC

SEP 20 1982

82 09 20 017

DTIC FILE COPY

ON THE ASYMPTOTIC DISTRIBUTION OF THE
SIZE OF A STOCHASTIC EPIDEMIC

by

THOMAS SELLKE

TECHNICAL REPORT NO. 17
MAY 1982

Prepared Under the Auspices
of
Office of Naval Research Contract
N00014-77-C-0306 (NR-042-373)

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

ON THE ASYMPTOTIC DISTRIBUTION OF
THE SIZE OF A STOCHASTIC EPIDEMIC

by

Thomas Sellke
Stanford University

ABSTRACT

For a stochastic epidemic of the type considered by Bailey [1] and Kendall [3], Daniels [2] showed that "when the threshold is large but the population size is much larger, the distribution of the number remaining uninfected in a large epidemic has approximately the Poisson form." A simple, intuitive proof is given for this result without use of Daniels' assumption that the original number of infectives is "small". The proof is based on a construction of the epidemic process which is more explicit than the usual description.

Key words: limiting Poisson distribution, stochastic epidemic



| | |
|--------------------|--|
| Accession For | |
| DTIC | <input checked="checked" type="checkbox"/> |
| TAB | <input type="checkbox"/> |
| Unannounced | <input type="checkbox"/> |
| Classification | |
| By | |
| Distribution/ | |
| Availability Codes | |
| Avail. and/or | |
| Special | |
| A | |

ON THE ASYMPTOTIC DISTRIBUTION OF THE SIZE OF A STOCHASTIC EPIDEMIC

1. Introduction

Consider a population which at time $t = 0$ consists of $X(0) = n$ healthy individuals and $Y(0) = m$ individuals with a contagious infection. An epidemic in such a population is often modeled by a continuous-time Markov process as follows. (See, for example, Bailey [1] or Kendall [3].) If $X(t)$ and $Y(t)$ are the numbers of healthy individuals and infectious individuals, respectively, present at time t , then the transition probabilities are given by

$$\begin{aligned} P\{(X(t+\delta), Y(t+\delta)) = (x', y') | (X(t), Y(t)) = (x, y)\} \\ &= xy\delta + o(\delta) \quad \text{for } (x', y') = (x-1, y+1) \\ &= \rho y\delta + o(\delta) \quad \text{for } (x', y') = (x, y-1) \\ &= 1 - xy\delta - \rho y\delta + o(\delta) \quad \text{for } (x', y') = (x, y) . \end{aligned}$$

The transitions listed represent the infection of a healthy individual, the removal of an infectious individual from the population, and "no change", respectively. All other possible transitions in $[t, t+\delta]$ are assumed to have collective probability $o(\delta)$. The positive constant ρ is variously called the threshold or the relative removal rate of the epidemic.

Note that states of the form $(x, 0)$ are absorbing, so that no more transitions occur after the last infectious individual has been removed. Absorption at $(x, 0)$ means that x individuals have escaped infection at the end of the epidemic and that $X(\infty) = x$, where $X(\infty) = \lim_{t \rightarrow \infty} X(t)$. It is easy to see that, with probability one, some such absorbing state is

reached eventually. Most of the work on this stochastic epidemic model has been directed toward finding the distribution of the number of new infections occurring during the course of the epidemic. This is equivalent to finding the distribution of the absorbing state. See Nagaev and Startsev [4] for an asymptotic analysis of limiting cases other than the one considered in this paper.

Under the assumption that $Y(0) = m$ is "small", Daniels [2] shows that "when the threshold is large but the population size is much larger, the distribution of the number remaining uninfected in a large epidemic has approximately the Poisson form with deterministic mean $ne^{-n/\rho}$." However, the asymptotic approximations used by Daniels in obtaining this result are not very enlightening, and he speculates that "there must be a direct argument in terms of the epidemic process itself to explain this Poissonlike behavior."

This paper gives a rather intuitive proof of Daniels' result. In addition, the assumption that m is small is dispensed with. The proof is based on a construction of the epidemic process which is more explicit than the usual description given above.

2. Construction of the Epidemic Process

Let the n originally healthy individuals be indexed by i , $1 \leq i \leq n$, and let the m originally infectious individuals be indexed by j , $1 \leq j \leq m$. Let $\{\hat{r}_j\}_{j=1}^m$ and $\{r_i\}_{i=1}^n$ be i.i.d. random variables with density $\rho e^{-\rho t}$ on $[0, \infty)$. Individual j in the original infectious group will remain infectious for \hat{r}_j time units before removal from the population. Individual i in the original healthy group will remain infectious for r_i time units if he becomes infected.

Let $\{l_i\}_{i=1}^n$ be i.i.d. random variables with density e^{-t} on $(0, \infty)$, independent of the \hat{r}_j 's and r_i 's. The variable l_i will be thought of as the "resistance to infection" of individual i in the original healthy group. Let $\{l_{(k)}\}_{k=1}^n$ be the associated order statistics, so that $l_{(1)} < l_{(2)} < \dots < l_{(n)}$. Let $r^{(k)} = r_i$ if $l_{(k)} = l_i$.

Now let the epidemic process proceed as follows. The originally infected individual j remains in the population for r_j time units, after which he is removed. The healthy individual i accumulates "exposure to infection" at a rate equal to the number of infected individuals present. When the total exposure to infection of healthy individual i reaches l_i , individual i becomes infected and then remains in the population for an additional r_i time units before removal.

It remains to be shown that the resulting process is Markov with the correct transition probabilities. Suppose $(X(t), Y(t)) \sim (x, y)$. The probability that a particular infected individual is removed in the time interval $[t, t+\delta]$ is $\rho\delta + o(\delta)$ because the distribution of the \hat{r}_j 's and r_i 's has constant hazard rate ρ . The probability that exactly one of the y infected individuals is removed in $[t, t+\delta]$ is therefore $\rho y\delta + o(\delta)$. The probability that a particular one of the x healthy individuals will become infected in $[t, t+\delta]$ is $y\delta + o(\delta)$, so that the probability that exactly one of the healthy individuals becomes infected is $xy\delta + o(\delta)$. It follows that the transition probabilities are as desired. The Markov property follows from the memoryless property of the exponential distribution.

Let v be the number of new infections occurring during the course of the epidemic. If $l_{(1)} > \sum_{j=1}^m \hat{r}_j$, then all originally infectious individuals are removed before the resistance to infection of any healthy individual has

been exceeded, so that $v = 0$. Otherwise, the originally healthy individual associated with $l_{(1)}$ becomes the first new infection, and $v \geq 1$. An easy induction argument shows that $v + 1$ is the smallest k , $1 \leq k \leq n$, for which

$$l_{(k)} > \sum_{j=1}^m \hat{r}_j + \sum_{i=1}^{k-1} r^{(i)}.$$

If this inequality does not hold for any k , $1 \leq k \leq n$, then $v = n$.

Define

$$R = \sum_{j=1}^m \hat{r}_j + \sum_{i=1}^v r^{(i)}.$$

Then R is the amount of "exposure to infection" withstood by those individuals who remain healthy at the end of the epidemic, and $X(\infty) = n - v$ is the number of l_i 's greater than R .

3. Statement of the Theorem and an Outline of the Proof

Consider a sequence of epidemics with parameters n_k , m_k , and ρ_k , $1 \leq k < \infty$. Let v_k be the number of new infections in a realization of the k -th epidemic, so that $X_k(\infty) = n_k - v_k$ is the number of individuals who escape infection.

Theorem. If $n_k \rightarrow \infty$, $\rho_k \rightarrow \infty$, and

$$n_k \exp\left\{-\frac{n_k + m_k}{\rho_k}\right\} \rightarrow b, \quad 0 < b < \infty,$$

then $X_k(\infty)$ converges in distribution to a Poisson random variable with mean b .

Suppressing the subscript k , we can summarize the proof as follows. The fact that ρ is $o(m+n)$ implies that, with high probability, all but a tiny fraction of the population becomes infected. Thus, R will be close to $\sum_{j=1}^m \hat{r}_j + \sum_{i=1}^n r_i$, which is in turn close to $(m+n)/\rho$ with high probability. If R is close enough to $(m+n)/\rho$, then $X(\infty)$ will equal the number of ℓ_i 's which are greater than $(m+n)/\rho$. The number of ℓ_i 's greater than $(m+n)/\rho$ has the distribution Binomial $(n, e^{-(m+n)/\rho})$, which converges in distribution to Poisson (b) .

4. Proof of the Theorem

The subscript k will again be suppressed.

Taking logarithms in $ne^{-(m+n)/\rho} \rightarrow b$ yields

$$\log n - \frac{m+n}{\rho} \rightarrow \log b ,$$

so that $\rho \sim \frac{m+n}{\log n}$. Thus, ρ is $o(m+n)$, but $(m+n)^\gamma$ is $o(\rho)$ for $0 < \gamma < 1$.

Lemma 1. Let $0 < \varepsilon < 1$ be given. If $\rho < \varepsilon^2(m+n)$, then

$$P\{X(\infty) > \varepsilon(m+n)\} < \left[\frac{\rho}{\varepsilon(m+n)}\right]^m < \varepsilon^m \leq \varepsilon .$$

Proof of Lemma 1.

As in Kendall [3], we can view the population of infected individuals as a continuous-time birth-and-death process with a variable birth-rate. The ratio of death-rate to birth-rate is $\rho/X(t)$, which is less than $\rho/(\varepsilon(m+n))$, until $X(t) \leq \varepsilon(m+n)$. The probability that a birth-and-death

process starting at m and with a death-rate to birth-rate ratio $q < 1$ is ever absorbed at zero is q^m .

Lemma 2. Let $0 < \epsilon < 1$. For n sufficiently large,

$$P\{R < \frac{(1-2\epsilon)(m+n)}{\rho}\} < 2\epsilon.$$

Proof of Lemma 2.

By Lemma 1, R is greater than the sum of the first $(1-\epsilon)(m+n)$ terms of $\{\hat{r}_1, \hat{r}_2, \dots, \hat{r}_m, r^{(1)}, r^{(2)}, \dots, r^{(n)}\}$ with probability greater than $(1-\epsilon)$ for sufficiently large n . The sum of the first $(1-\epsilon)(m+n)$ terms has mean $(1-\epsilon)(m+n)/\rho$ and variance $(1-\epsilon)(m+n)/\rho^2$. The Chebyshev inequality now implies the lemma.

Lemma 3. Let $0 < \epsilon < 1$. For n sufficiently large,

$$P\{X(\infty) > n^{3\epsilon}\} < 3\epsilon.$$

Proof of Lemma 3.

The number of \hat{r}_i 's which are greater than $(1-2\epsilon)(m+n)/\rho$ is Binomial $(n, e^{-(1-2\epsilon)(m+n)/\rho})$. This distribution has mean

$$ne^{-(1-2\epsilon)(m+n)/\rho} \sim n\left(\frac{b}{n}\right)^{(1-2\epsilon)} = b^{1-2\epsilon} n^{2\epsilon}.$$

Application of Lemma 2 and the Markov inequality finishes the proof of Lemma 3.

From Lemma 3, it is easy to see that $X(\infty)$ is $o(\sqrt{n})$ in probability. Thus, except on a set of small probability, R is greater than the sum of

the first $m + n - \sqrt{n}$ terms of $\{\hat{r}_1, \hat{r}_2, \dots, \hat{r}_m, r^{(1)}, r^{(2)}, \dots, r^{(n)}\}$.

R is of course less than or equal to the sum of all the terms. An argument like the proof of Lemma 2 shows that, with probability approaching one,

$$\frac{m+n}{\rho} - \frac{(m+n)^{2/3}}{\rho} < R < \frac{m+n}{\rho} + \frac{(m+n)^{2/3}}{\rho}.$$

The number of ℓ_1 's greater than

$$\frac{m+n}{\rho} \pm \frac{(m+n)^{2/3}}{\rho}$$

is distributed as a Binomial $(n, \exp\{-\frac{m+n}{\rho} \mp \frac{(m+n)^{2/3}}{\rho}\})$, which has mean

$$n \exp\{-\frac{m+n}{\rho} \mp \frac{(m+n)^{2/3}}{\rho}\} \sim b \exp\{\mp \frac{(m+n)^{2/3}}{\rho}\} + b.$$

Thus, with probability approaching one, $X(\infty)$ is less than a

$$\text{Binomial}(n, \exp\{-\frac{m+n}{\rho} + \frac{(m+n)^{2/3}}{\rho}\}) \text{ random variable}$$

and greater than

$$\text{Binomial}(n, \exp\{-\frac{m+n}{\rho} - \frac{(m+n)^{2/3}}{\rho}\}) \text{ random variable}.$$

Since both of these distributions converge in law to a Poisson with mean b , the theorem is proved.

References

- [1] Bailey, N.T.J., The Mathematical Theory of Infectious Diseases and Its Applications, London: Griffin, 1975.
- [2] Daniels, H.E., "The Distribution of the Total Size of an Epidemic," Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 4, pp. 281-293.
- [3] Kendall, D.G., "Deterministic and Stochastic Epidemics in ~~eed~~ Populations," Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Vol. 4, 149-165.
- [4] Nagaev, A.V. and Startsev, A.N., "The Asymptotic Analysis of a Stochastic Model of an Epidemic," Theory of Probability and its Applications, 15, (1970), pp. 98-107.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

| REPORT DOCUMENTATION PAGE | | READ INSTRUCTIONS BEFORE COMPLETING FORM |
|--|-----------------------------------|--|
| 1. REPORT NUMBER 17 | 2. GOVT ACCESSION NO. 1119 370 | 3. RECIPIENT'S CATALOG NUMBER |
| 4. TITLE (and Subtitle) ON THE ASYMPTOTIC DISTRIBUTION OF THE SIZE OF A STOCHASTIC EPIDEMIC | | 5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT |
| 7. AUTHOR(s) THOMAS SELLKE | | 6. PERFORMING ORG. REPORT NUMBER |
| 9. PERFORMING ORGANIZATION NAME AND ADDRESS DEPT. OF STATISTICS STANFORD UNIVERSITY STANFORD, CALIF. | | 8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0306 |
| 11. CONTROLLING OFFICE NAME AND ADDRESS STATISTICS & PROBABILITY PROGRAM CODE 436 OFFICE OF NAVAL RESEARCH ARLINGTON VA 22217 | | 10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-373 |
| 14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) | | 12. REPORT DATE MAY 1982 |
| | | 13. NUMBER OF PAGES 8 |
| | | 15. SECURITY CLASS. (of this report) UNCLASSIFIED |
| | | 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE |
| 16. DISTRIBUTION STATEMENT (of this Report) APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED | | |
| 17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report) | | |
| 18. SUPPLEMENTARY NOTES | | |
| 19. KEY WORDS (Continue on reverse side if necessary and identify by block number) LIMITING POISSON DISTRIBUTION STOCHASTIC EPIDEMIC | | |
| 20. ABSTRACT (Continue on reverse side if necessary and identify by block number) PLEASE SEE OTHER SIDE | | |

DD FORM 1 JAN 73 1473

EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6001

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

**ON THE ASYMPTOTIC DISTRIBUTION OF
THE SIZE OF A STOCHASTIC EPIDEMIC**

by

**Thomas Sellke
Stanford University**

ABSTRACT

For a stochastic epidemic of the type considered by Bailey [1] and Kendall [3], Daniels [2] showed that "when the threshold is large but the population size is much larger, the distribution of the number remaining uninfected in a large epidemic has approximately the Poisson form." A simple, intuitive proof is given for this result without use of Daniels' assumption that the original number of infectives is "small". The proof is based on a construction of the epidemic process which is more explicit than the usual description.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

